

SCRAMBLING, DESCRAMBLING AND SECURED DISTRIBUTION OF
AUDIOVISUAL SEQUENCES STEMMING FROM VIDEO CODERS
BASED ON A PROCESSING BY WAVELETS

[0001] The present invention is relative to the area of the processing of video sequences encoded with the aid of video coders based on wavelet technology.

[0002] The present invention proposes supplying a process and a system that permit the visual scrambling of a video sequence and the subsequent recomposing (descrambling) of its original content from a digital video stream obtained by an encoding based on a transform in [into] wavelets.

[0003] The present invention is relative in particular to a device capable of securely transmitting a set of video streams with a high visual quality to a viewing screen of the television screen type and/or for being recorded on the hard disk or on any other recording support of a box connecting the telecommunication network to a viewing screen such as a television screen or a personal computer monitor while preserving the audiovisual quality but avoiding any fraudulent use such as the possibility of making pirated copies of films or audiovisual programs recorded on the hard disk or any other recording support of the decoder box. The invention also concerns a client-server system, in which the server supplies the stream permitting the viewing of the secured distribution video film and the client reads and displays the digital audiovisual stream.

[0004] It is possible with the current solutions to transmit films and audiovisual programs in digital form via broadcasting networks of the microwave [herzian], cable, satellite type, etc. or via telecommunication networks of the DSL (digital subscriber line) type or BLR (local radio loop) type or via DAB (digital audio broadcasting) networks, etc. Moreover, in order to avoid

the pirating of works broadcast in this manner, the latter are frequently encrypted or scrambled by various means well known to an expert in the art.

[0005] Concerning the processing of video sequences encoded with wavelet technology, the prior art contains US patent 6, 370, 197 entitled “Video Compression Scheme Using Wavelets” in which the authors detail a method of coding a video sequence based on a wavelet transform and generating a nested digital stream. This prior art does not propose any method for protecting the stream and/or scrambling the video sequence.

[0006] EP patent 0734164 is also known and presents a process and a device for increasing the efficacy of coding brought about by video encoders based on the classified vectorial quantification by optimizing the coding in such a manner as to not have to transmit the classification information in the encoded binary stream. This prior art applies to video streams stemming from a DCT transform or a wavelet transform. To this end the entering video signal is divided into a plurality of subbands, e.g., the DC coefficients are arranged in one subband and the AC coefficients in remaining subbands, followed by a formatting in blocks of identical size, each block of which includes a DC coefficient and a multitude of AC coefficients. A selection signal is then generated, representing the vectorial quantification class corresponding to each assembled block. This stage is followed by classification for the vectorial quantification by the generation of parameters relative to the evolution of the DC coefficients in the horizontal and the vertical direction, and by the differential entropic encoding of the DC coefficients relative to the assembled blocks for generating a first encoded video signal. The AC coefficients are classified and encoded separately with the aid of an entropic encoding as a function of the selection information for generating a second encoded video signal. The two signals generated in this manner are formatted for the transmission. For the decoding, no classification information is

transmitted – it is reconstructed from the DC coefficients encoded and transmitted to the decoder. This solution concerns a process relative to the digital compression and the encoding of video streams stemming from the DCT transform and the wavelet transform. The description of the process indicates the stages to be applied for implementing a classified vectorial quantification that increases the compression and the effectiveness of the encoding. A single stream is transmitted to the receiver. The technical problem and the objective posed in this document are to optimize the digital format and it has the task of obtaining a digital stream formatted at the output of a digital encoder. The process described in this solution does not permit a securing of the video stream and does not offer any protection against illicit uses of video streams stemming from the encoding in wavelets.

[0007] As concerns the protection of images coded in wavelets, the prior art contains document EP 1 033 880, that is relative to a process and a device for protection by modifications applied to the spatial-frequency coefficients. These modifications are of the type: Modification of the sign bits of the coefficients, modification of the improvement bits of the coefficients, the choice of the appropriate coefficients belonging to a frequency subband for shifting (exchanging) them, rotation of a block regrouping the frequency coefficients arranged in increasing order while attempting to respect to the maximum the static properties and the entropy of the original signal. Each type of modification is conditioned with the aid of a key. The data protected in this manner is then passed through an entropic encoder and a bitstream in conformity with the norm is generated. This prior art represents a solution of encrypting with the aid of keys and as a consequence a single stream is transmitted to the receiver and all the elements constituting the original stream are located within the protected stream. This document concerns a solution that does not respond in a satisfactory manner to the protection of the transmitted video stream.

Moreover, as a consequence of the modifications before the entropic encoding the statistical properties are modified and the size of the stream and the transmission rate increase. Consequently, this prior art does not satisfy the objectives of high security that guarantee a process without loss, the subject matter of the present invention.

[0008] Another reference from the prior art is document WO 00 31964 A relative to a method and equipment for the partial encrypting of images in order to protect them and to optimize the storage location. A first part of the image is compressed to a low quality without encryption and a second part of the image is encrypted. When the first and the second part are reunited the image is obtained with maximal quality. The second part is encrypted and comprises two sections encrypted in different manners. The decryption of the first section and its combination with the first part restores the initial image with an average quality. The decryption of the second section, its combination with the first section and the first part restores the original image with maximal quality. The image can also be partitioned into multiple independent sections, each section of which is encrypted with its own method and its own key. The protection method in this prior art is the encryption and consequently all the original elements of the stream remain within the protected stream and the restoration of the entire content from only the protected stream is possible in the instance that an ill-disposed person finds or simulates the encryption keys. As was the case for the preceding document, this solution does not furnish satisfactory security against the pirating of the video stream. Also, the size of the protected stream is different from the size of the original stream. This prior art therefore does not resolve the problem of high security while procuring a fine granularity in the quality of the reconstituted video sequences processed in the present invention.

[0009] In the prior art concerning the secured distribution of audiovisual streams organized in multilayers based on the client-server principal, US patent 2001/0053222 proposes a process and a system for the protection of video streams encoded according to the MPEG-4 norm. The audiovisual stream is composed of several audio and video objects managed by a scenic composition. One of the objects of the video stream is encrypted with the aid of a key generated in four encryption stages that is periodically renewed. The protected objects are video objects. The encrypted object is multiplexed with the other objects and the entire stream is sent to the user. The MPEG-4 stream is recomposed in the addressed equipment by the decryption module, that reconstitutes the original video stream from the encrypted video stream and by regenerating the encryption key from encryption information previously sent and from information contained in the encrypted stream. Given that the entire protected content of the video objects is located in the stream sent to the user, an ill-intentioned person who finds the encryption keys would be able to decrypt this protected content and to view or broadcast it. This prior art therefore does not entirely resolve the problem of securing the video stream.

[0010] Contrary to the majority of these “classic” protection systems, the process in conformity with the invention ensures a high level of protection while reducing the volume of information necessary in order to have access to the original content from the protected content.

[0011] The protection, realized in a manner in conformity with the invention, is based on the principle of the deletion [removal] and replacement of certain information coding the original visual signal by any method, e.g.: Substitution, modification, permutation [swapping, shifting] or movement of the information. This protection is also based on a knowledge of the structure of the binary stream at the output of the wavelet-based visual encoder.

[0012] The present invention relates to the general principle of a process for securing an audiovisual stream. The objective is to authorize video services on demand and a la carte via all the broadcasting networks and the local recording in the digital decoder box of the user, as well as the direct viewing of television channels. The solution consists in extracting and permanently preserving outside of the user's dwelling, and in fact in the broadcasting and transmitting network, a part of the digital audiovisual stream recorded at the client's or directly broadcasted, which part is of primary importance for viewing said digital audiovisual stream on a television screen or monitor type screen, but which has a very small volume relative to the total volume of the digital audiovisual stream recorded at the user's or received in real time. The lacking part will be transmitted via the broadcasting or transmitting network at the moment of the viewing of said digital audiovisual stream.

[0013] Once the digital audiovisual stream has been modified and separated into two parts the larger part of the modified audiovisual stream, called "modified main stream" will therefore be transmitted via a classic broadcasting network whereas the remaining part, called "complementary information" will be sent on demand via a narrow-band telecommunication network such as the classic telephone networks or cellular networks of the GSM, GPRS or BLR types, or also by using a subset of the bandwidth shared on a cable network. The original digital audio stream is reconstituted in the addressed equipment (decoder) by a synthesis module from the modified main stream and the complementary information.

[0014] The invention realizes a protection system comprising an analysis - scrambling and descrambling module based on a digital format stemming from the encoding of a video stream based on wavelet transforms. The analysis and scrambling module proposed by the invention is based on substitution by "decoys" or the modification of part of the coefficients stemming from

the transformation in wavelets. The fact of having removed and substituted part of the original data of the original digital stream during the generation of the modified main stream does not allow the restoration of said original stream only from the data of this modified main stream. Several variants of the scrambling and descrambling process are implemented and illustrated with exemplary embodiments by the characteristics of “scalability” of the wavelet transform, which notion of “scalability” is defined from the English expression “scalability,” that characterizes an encoder capable of encoding or a decoder capable of decoding an ordered set of digital streams in such a manner as to produce or reconstitute a multilayer sequence.

[0015] The invention concerns in its most general meaning a process for the secured distribution of video sequences according to a digital stream format stemming from an encoding based on a processing by wavelets, constituted by frames comprising blocks containing coefficients of wavelets describing the visual elements, characterized in that an analysis of the stream is made prior to the transmission to the client equipment in order to generate a modified main stream by deletion and replacement of certain information coding the original stream and presenting the format of the original stream, and complementary information of any format comprising this digital information coding the original stream and suitable for permitting the reconstruction of these modified frames, then this modified main stream and this complementary information generated in this manner are transmitted separately from the server to the addressed equipment.

[0016] The protection is brought about by the deletion of the original elements and by substituting them with decoys, which original extracted elements are stored separately in the complementary information. The fact of having removed and substituted a part of the original

data of the original video stream during the generation of the modified main stream does not allow the restoration of the original stream from only the data of this modified main stream.

[0017] The video stream is entirely protected (all the subbands) and entirely transmitted via the network or via a physical support to the user, independently of his rights. The partial restoration is carried out via the sending of part of the complementary information containing the original elements either directly or in a progressive mode.

[0018] The analysis and scrambling module decides how to visually degrade the video stream as a function of its structure and its properties of scalability resulting from the transform in wavelets. The study concerns the impact of the modification of different parts of the stream (coefficients, subbands, layers of scalability, zones of interest) on the visual degradation.

[0019] The scrambling is preferably carried out by modifying the wavelet coefficients belonging to at least one temporal [time-division] subband resulting from the temporal analysis.

[0020] The scrambling is advantageously brought about by modifying coefficients of wavelets belonging to at least one spatial subband resulting from the spatial analysis of a temporal subband.

[0021] The scrambling is advantageously brought about by modifying coefficients of wavelets belonging to at least one temporal subband resulting from a temporal analysis of one spatial subband.

[0022] The wavelet coefficients to be modified are advantageously selected according to laws that are random and/or defined a priori [beforehand].

[0023] According to a particular embodiment the parameters for the scrambling are a function of the properties of temporal scalability and/or of spatial scalability and/or of qualitative

scalability and/or of transmission rate scalability and/or of scalability by regions of interest offered by the digital streams generated by the wavelet-based coders.

[0024] The visual intensity of the degradation of the video sequences obtained is advantageously determined by the quantity of modified wavelet coefficients in each spatial-temporal subband.

[0025] The intensity of the visual degradation of the video sequences decoded from the modified main stream is advantageously a function of the position in the original digital stream of the modified data, which data represents, according to its positions, the values quantified according to different precisions [accuracies] of the wavelet coefficients belonging to a spatial-temporal subband.

[0026] The intensity of the visual degradation of the video sequences decoded from the modified main stream is advantageously determined according to which quality layer of the modified wavelet coefficients they belong to in each spatial-temporal subband.

[0027] According to a particular embodiment the modification of the wavelet coefficients is carried out directly in the binary stream.

[0028] According to a variant the modification of the wavelet coefficients is carried out with a partial decoding.

[0029] According to another variant the modification of the wavelet coefficients is carried out during the coding or by carrying out a decoding then a complete re-encoding.

[0030] The size of the modified main stream is advantageously strictly identical to the size of the original digital video stream.

[0031] According to another variant the substitution of the wavelet coefficients is carried out with random or calculated values.

[0032] The duration of the visual scrambling obtained in a group of frames is preferably determined as a function of the temporal subband to which the modified wavelet coefficients belong.

[0033] The visual scrambling obtained in a group of frames is advantageously limited spatially in a region of interest of each frame.

[0034] In addition, the complementary information is organized in layers of temporal and/or spatial and/or qualitative and/or transmission rate scalability and/or scalability by region of interest.

[0035] In one variant the stream is progressively descrambled with different layers of quality and/or resolution and/or frame rate and/or according to a region of interest via the sending of certain parts of the complementary information corresponding to the layers of qualitative and/or spatial and/or temporal scalability and/or scalability for a region of interest.

[0036] According to another variant the stream is partially descrambled according to different levels of quality and/or resolution and/or frame rate and/or according to a region of interest via the sending of a part of the complementary information corresponding to the layer or layers of qualitative and/or spatial and/or temporal scalability and/or scalability for this region of interest.

[0037] According to a particular embodiment a synthesis of a digital stream in the original format is calculated in the addressed equipment as a function of this modified main stream and of this complementary information.

[0038] According to a particular embodiment the transmission of this modified main stream is realized via a physically distributed material support (CD-ROM, DVD, hard disk, flash memory card).

[0039] The modified main stream advantageously undergoes operations of transcoding, of rearrangement and/or of the extraction of frames or of groups of frames during its transmission.

[0040] The transmission of this complementary information is advantageously realized via a physically distributed support material (flash memory card, smart card).

[0041] The modification of the wavelet coefficients is preferably perfectly reversible (lossless process) and the digital stream reconstituted from the modified main stream and from the complementary information is strictly identical to the original stream.

[0042] The modification of the wavelet coefficients is advantageously perfectly reversible (lossless process) and the portion of the digital stream reconstituted from the modified main stream and from the complementary information is strictly identical to the corresponding portion in the original stream.

[0043] According to a particular variant the reconstitution of a descrambled video stream is controlled and/or limited in terms of predefined frame rate and/or resolution and/or transmission rate and/or quality as a function of the user rights.

[0044] According to another variant the reconstitution of a descrambled video stream is limited in terms of frame rate and/or resolution and/or transmission rate and/or quality as a function of the viewing apparatus on which it is visualized.

[0045] According to another variant the reconstitution the descrambled video stream is carried out in a progressive manner in stages up to the reconstitution of the original video stream.

[0046] The invention also relates to a system for the fabrication of a video stream, comprising at least one multimedia server containing the original video sequences and comprising a device for analyzing the video stream, a device for separating the original video stream into a modified main stream by deletion and replacement of certain information coding

the original visual signal and into complementary information as a function of this analysis, and at least one device in the addressed equipment for the reconstruction of the video stream as a function of this modified main stream and of this complementary information.

[0047] The present invention will be better understood from a reading of the following description of a non-limiting exemplary embodiment that refers to the figure describing the total architecture of a system for implementing the process of the invention.

[0048] The described protection of the visual streams is worked out based on the structure of the binary streams and their characteristics due to the encoding based on wavelets. This structure will be recalled in the following.

[0049] A video coder based on a processing by wavelets realizes a temporal and spatial decomposition of an initial video sequence in order to obtain a set of coefficients of spatial-temporal wavelets. These coefficients are then quantified, then coded by an entropic coder in order to generate one or several nested binary streams possessing properties of temporal scalability and/or spatial (or resolution) scalability and/or qualitative scalability and/or transmission rate scalability and/or scalability of regions of interest.

[0050] The property of temporal scalability is the possibility of decoding from a single or several nested binary streams video sequences of which the display frequency of the frames (number of frames per second) is variable.

[0051] Literature well-known to an expert in the art employs the notions in English of “frames” (= Fr. “trames”) and of “frame rate” (number of frames per second), which notions will be used in the following for describing the present invention.

[0052] The property of spatial scalability is the possibility of decoding from a single or several nested binary streams video sequences of which the spatial (size) resolution of the frames is variable.

[0053] The property of qualitative scalability is the possibility of decoding from a single or several nested binary streams video sequences of which the visual quality of the frames, measured according to objective and/or subjective criteria, is variable.

[0054] The property of transmission rate scalability is the possibility of decoding from a single or several nested binary streams video sequences according to an average transmission rate (average number of information bits per second).

[0055] The property of scalability by region of interest is the possibility of decoding from a single or several nested binary streams one or several targeted zones in the video sequence.

[0056] During the encoding an original video sequence is segmented into groups of N successive frames called GOF (group of frames), and each GOF is then processed in an independent manner during the encoding. Note a GOF with length N GOF=(F_0, F_1, \dots, F_{N-1}), F_i [sic] being the frames for $i = 0, 1, 2, \dots, N-1$. The spatial-temporal wavelet coefficients are generated in two successive stages and in accordance with a spatial and temporal analysis of the frames of the GOF.

[0057] The first stage consists in performing a temporal analysis of the N frames of each GOF in accordance with the estimated direction of the movement (temporal analysis with estimation of movement) in order to remove the temporal redundancies and to spatially concentrate the energy and the information due to the movement in the frames stemming from the temporal analysis. This temporal analysis can be performed according to different spatial resolutions and after the decomposition into wavelets of each frame of the GOF. The estimation

of movement is performed independently within each spatial subband (multi-resolution estimation of movement).

[0058] The second stage consists in performing a spatial analysis of the N frames resulting from the temporal analysis with the aid of a decomposition into wavelets in order to remove the spatial redundancies and to concentrate the energy due to spatial discontinuities present in each frame.

[0059] The temporal analysis is performed in several iterations.

[0060] In the first iteration p successive frames of the original GOF are analyzed with predefined wavelet filters f_L and f_H with length p and after estimation and compensation of the movement for each frame and in relation to one or several frames called reference frames. Generally, $p=2$ and two successive frames (F_{2i}, F_{2i+1}) , $i=0, \dots, N/2$ are filtered and engender a frame called “low-frequency” or “average” and noted [written] $L_i=f_L(F_{2i}, F_{2i+1})$ and a frame called “high-frequency” and or “difference” and noted $H_i=f_H(F_{2i}, F_{2i+1})$. Thus, at the first iteration of the temporal analysis stage a subset $t-L_1$ of frames of type L with length $N/2$ and a subset $t-H_1$ of frames of type H with length $N/2$ are generated such as

$$t-L_1=(L_0, L_1, \dots, L_{N/2}),$$

$$t-H_1=(H_0, H_1, \dots, H_{N/2}).$$

[0061] At each following iteration $k>1$ the estimation/compensation of movement and the temporal filtering are iterated for the subset of frames $t-L_{k-1}$ of type L obtained in the iteration $k-1$ and two new subsets of frames $t-L_k$ and $t-H_k$ are generated whose length (number of frames) is reduced by a factor of 2 relative to $t-L_{k-1}$. In certain instances the iteration also relates to the subset of frames $t-H_{k-1}$.

[0062] The total number of iterations during the temporal analysis is noted n_T . It is comprised between 1 and $N/2$. n_T+1 temporal subbands are generated at the end of the temporal analysis.

[0063] For example, with $N=16$ and $n_T=4$:

$$L_i^k = f_L(L_{2i}^{k-1}, L_{2i+1}^{k-1}),$$

$$H_i^k = f_H(L_{2i}^{k-1}, L_{2i+1}^{k-1})$$

GOF = (F_0, F_1, \dots, F_{15})	
$t-L_1 = (L_0, \dots, L_7)$	$t-H_1 = (H_0, \dots, H_7)$
$t-L_2 = (LL_0, \dots, LL_3)$	$t-H_2 = (LH_0, \dots, LH_3)$
$t-L_3 = (LLL_0, LLL_1)$	$t-H_3 = (LLH_0, LLH_1)$
$t-L_4 = (LLLL_0)$	$t-H_4 = (LLLH_0)$

[0064] The temporal analysis of a GOF with length $N=16$ with $n_T=4$ therefore generates 16 new frames divided into $n_T+1=4+1=5$ temporal subbands:

- Subband $t-L_4$: 1 frame of type L: $LLLL_0$,
- Subband $t-H_4$: 1 frame of type H: $LLLH_0$,
- Subband $t-H_3$: 2 frames of type H: LLH_0, LLH_1 ,
- Subband $t-H_2$: 4 frames of type H: LH_0, \dots, LH_3 ,
- Subband $t-H_1$: 8 frames of type H: H_0, \dots, H_7 .

[0065] The spatial analysis is then performed on each of the frames belonging to each temporal subband $t-L_i$ and $t-H_i$: Each frame is decomposed with the aid of a wavelet transform discrete at D levels, thus generating $3xD+1$ spatial subbands of wavelet coefficients for each frame. These spatial subbands are noted $s-LL_0, s-HL_1, s-LH_1, s-HH_1, s-HL_2, s-LH_2, s-HH_2, \dots, s-HL_D, s-LH_D, s-HH_D$.

[0066] At the end of the temporal and spatial analyses $(n_T+1) \times (3D+1)$ spatial-temporal subbands of wavelet coefficients are available:

- $t-L_{nT}(s-LL_0), t-L_{nT}(s-HL_1), t-L_{nT}(s-LH_1), t-L_{nT}(s-HH_1), \dots, t-L_{nT}(s-HL_D), t-L_{nT}(s-LH_D), t-L_{nT}(s-HH_D),$
- $t-H_{nT}(s-LL_0), t-H_{nT}(s-HL_1), t-H_{nT}(s-LH_1), t-H_{nT}(s-HH_1), \dots, t-H_{nT}(s-HL_D), t-H_{nT}(s-LH_D), t-H_{nT}(s-HH_D),$
- \dots

$t-H_1(s-LL_0), t-H_1(s-HL_1), t-H_1(s-LH_1), t-H_1(s-HH_1), \dots, t-H_1(s-HL_D), t-H_1(s-LH_D), t-H_1(s-HH_D).$

[0067] The wavelet coefficients of each spatial-temporal subbands are then compressed progressively by bit plane with the aid of an entropic coder with the task of removing the statistical redundancies existing in a fixed set of wavelet coefficients. The entropic coder generates a binary stream for each set of independently coded wavelet coefficients which binary stream can be sectioned into several substreams divided according to different quality layers.

[0068] After an analysis of the structure in subbands previously described, the analysis and scrambling module in conformity with the invention performs modifications (by permutation and/or substitution and/or thresholding) of a subset of the wavelet coefficients belonging to one or several spatial-temporal subbands. These modifications introduce a visually perceptible degradation (scrambling) of the video sequence decoded from these modified coefficients. A control [verification, check] of the spatial and/or temporal extent and/or according to the layers of quality of the scrambling as well as a control of the intensity of the degradation due to the scrambling are possible as a function of the number of modified coefficients, their localizations in a spatial subband, their belonging to the spatial-temporal subbands, their belonging to one or several quality layers, their position in the set of coefficients belonging to a single spatial-temporal subband and of the type of modification.

[0069] The N/l_x consecutive frames of the GOF are scrambled by modifying the wavelet coefficients in spatial subbands of a temporal subband $t\text{-}X$ with length l_x (i.e., containing l_x frames).

[0070] The selection of the type of spatial subband to which the wavelet coefficients (s-HL or s-LH or s-HH) belong permits a control of the visual aspect of the scrambling: For the s-HL subband artifacts of vertical direction appear on the frames (degradation of the vertical spatial discontinuities), for the s-LH subband horizontal artifacts appear (degradation of the horizontal spatial discontinuities) and for the s-HH subband artifacts of the “checkerboard” type appear (conjoined degradations of the horizontal and vertical spatial discontinuities).

[0071] The selection of the level of resolution r to which the spatial subband ($s\text{-}LL_r$ or $s\text{-}HL_r$ or $s\text{-}LH_r$ or $s\text{-}HH_r$) belongs permits a control of the spatial extent of the scrambling engendered by the modification of the wavelet coefficients : The closer r is to 0, the greater the spatial extent.

[0072] A modification of the wavelet coefficients belonging to a subband with resolution $r \geq 0$ generates a scrambling that is visible on all the frames decoded with spatial resolutions greater than $r+1, r+2, \dots, R$.

[0073] A modification of the wavelet coefficient belonging to a quality layer q generates a scrambling that is visible on all the decoded frames considering at least the q first quality layers.

[0074] The modification of the spatial-temporal wavelet coefficients is performed after a partial decoding of the binary stream generated in accordance with a standard or norm or an algorithm or an encoding format. Once the modification has been made, a re-encoding of the coefficients is performed in order to generate a binary stream with the identical size that respects

the conformity relative to the standard or norm or algorithm or encoding format that generated the original binary stream.

[0075] Subunits of bits inside the original binary stream representing the coded spatial-temporal wavelet coefficients are modified without decoding and without disturbing the conformity of the stream relative to the standard or norm or algorithm or encoding format that generated the original binary stream.

[0076] The selection of the spatial-temporal wavelet coefficients to be modified in a spatial-temporal subband is made in a manner that is random and/or in accordance with previously defined rules.

[0077] The modified main stream advantageously has a size identical to that of the original video stream.

[0078] The scrambling generated in this manner has properties of temporal, spatial, qualitative and transmission rate scalability and scalability by zone of interest.

[0079] The complementary information relative to this scrambling generated in this manner is advantageously organized in layers of temporal, spatial, qualitative and transmission rate scalability and scalability by zone of interest.

[0080] The scrambling has, as a function of the number of GOF and/or of the number of frames scrambled in a GOF, a temporal scalability comprised between: “All the frames of all the GOFs (maximal scrambling)” and “no frame of any GOF” (non-scrambled sequence).

[0081] The scrambling has, as a function of the resolutions of the spatial subbands to which the modified wavelet coefficients belong, a spatial scalability comprised between: “All the resolutions are scrambled” (i.e., from resolution $r=0$ to resolution $r=R$) and “none of the resolutions is scrambled.”

[0082] The scrambling has, as a function of the number of modified wavelet coefficients and of the resolutions of the spatial subbands to which they belong, a qualitative scalability ranging from: “The entirety of each frame is scrambled,” “certain spatial regions of each frame are scrambled” (regions of interest) and “no scrambling was applied to the frames.”

[0083] In a reciprocal manner, the descrambling also has the different scalabilities stated (temporal, spatial, qualitative and transmission rate and by zone of interest).

[0084] This descrambling advantageously permits the different scalabilities stated (temporal, spatial, qualitative and transmission rate and by zone of interest) to be addressed by virtue of the sending of certain parts of the complementary information corresponding to different layers of scalability (temporal, spatial, qualitative and transmission rate and by zone of interest), thus giving access to different level of quality/resolution/frame rate for the video sequence decoded from the partially descrambled stream.

[0085] The different levels of quality/resolution/frame rate of the video sequence are advantageously obtained from the partially descrambled stream via the sending of a part of the complementary information by layer of scalability (temporal, spatial, qualitative and transmission rate and by zone of interest).

[0086] The principle of scrambling and of descrambling based on these different scalabilities will be better understood with the aid of the following preferred, non-limiting exemplary embodiment.

[0087] In the attached drawing the figure represents a particular preferred embodiment of the client-server system in conformity with the invention.

[0088] The original stream 1 is directly in digital form or in analog form. In this latter instance the analog stream is converted by a wavelet-based coder (not shown) into a digital

format 2. The video stream to be secured 2 is passed to analysis and scrambling module 3 that will generate a modified main stream 5 in the identical format as input stream 2 aside from the fact that certain coefficients were replaced by values different than the original ones and is stored in server 6. Complementary information 4 of any format is also placed in server 6 and contains information relative to the elements of the images that were modified, replaced, substituted or moved and to their value or location in the original stream.

[0089] Stream 5 with a format identical to the original stream is then transmitted via high line speed network 9 of the microwave [hertzian, cable, satellite type, etc. to the terminal of spectator 8 and more precisely onto his hard disk 10. When spectator 8 request to view the film present on his hard disk 10, two things are possible: Either spectator 8 does not have all the rights necessary for viewing the film in which case video stream 5 generated by scrambling module 3 present on hard disk 10 is passed to synthesis system 13 via reading buffer memory 11 that does not modify it and transmits it identically to a display reader capable of decoding it 14, and its content, degraded visually by scrambling module 3, is displayed on viewing screen 15. Video stream 5 generated by scrambling module 3 is passed directly via network 9 to reading buffer memory 11 and then to synthesis system 13.

[0090] Video stream 5 advantageously undergoes a series of operations of transcoding and of rearrangement of its frames or groups of frames in network 9.

[0091] Or, the server decides that spectator 8 has the right to correctly view the film. In this instance synthesis module 13 makes a viewing request to server 6 containing the complementary information necessary 4 for the reconstitution of original video 2. Server 6 then sends complementary information 4 via telecommunication network 7 of an analog or digital telephone line type, DSL (digital subscriber line) or BLR (local radio loop), via DAB (digital audio

broadcasting) networks or via digital mobile telecommunication networks (GSM, GPRS, UMTS), which complementary information 4 permits the reconstitution of the original video in such a manner that spectator 8 can store it in a buffer memory 12. Synthesis module 13 then proceeds to a restoration of the scrambled video stream that it reads in its reading buffer memory 11, and modified fields whose positions it knows as well as the original values are restored by virtue of the content of the complementary information read in descrambling buffer memory 12. The quantity of information contained in complementary information 4 that is sent to the descrambling module is specific, adaptive and progressive for each spectator and depends on his rights, e.g., single or multiple use, right to make one or more private copies, delayed payment or payment in advance. The level (quality, quantity, type) of complementary information is also determined as a function of the visual quality required by the user. The wave-based video coding characterized by the previously described scalabilities permits the restoration of the video stream with levels of quality, resolution and frequency of different frames.

[0092] Modified main stream 5 is advantageously passed directly via network 9 to reading buffer memory 11, then to synthesis module 13.

[0093] Modified main stream 5 is advantageously inscribed (recorded) on a physical support such as a CD-ROM or DVD, hard disk, flash memory card, etc. (9bis). Modified main stream 5 is then read from physical support 9bis by disk reader 10bis of box 8 in order to be transmitted to reading buffer memory 11, then to synthesis module 13.

[0094] Complementary information 4 is advantageously recorded on a physical support 7bis with a credit card format constituted by a smart card or a flash memory card. This card 7bis will be read by module 12 of device 8 comprising card reader 7ter.

[0095] Card 7bis advantageously contains the applications and the algorithms that will be executed by synthesis system 13.

[0096] Device 8 is advantageously an autonomous, portable and mobile system.

[0097] The functioning of analysis and scrambling module 3 illustrating the selection of the scrambling performed will now be described in detail. The original video sequence is segmented into GOF with $N=16$ frames. The temporal analysis with $n_T=4$ iterations generates $n_T+1=5$ temporal subbands respectively processing:

- Subband t-L₄: 1 frame of type L: LLLL₀,
- Subband t-H₄: 1 frame of type H: LLLH₀,
- Subband t-H₃: 2 frames of type H: LLH₀, LLH₁,
- Subband t-H₂: 4 frames of type H: LH₀, LH₁, LH₂, LH₃,
- Subband t-H₁: 8 frames of type H: H₀, H₁, H₃, H₄, H₅, H₆, H₇.

[0098] The decomposition into five temporal subbands offers the possibility of restoring the initial video sequence according to five different frame rates. Each frame of resolution R in each temporal subband t-X is then decomposed spatially by a wavelet transform at D=4 levels, which yields the possibility of reconstituting the image with five different resolutions, thus generating for each $3 \times D + 1 = 13$ spatial subbands: LL₀, LH₁, HL₁, HH₁, LH₂, HL₂, HH₂, LH₃, HL₃, HH₃, LH₄, HL₄, HH₄.

[0099] As a consequence of such an encoding, the video sequence can therefore be decoded according to frame rates from $1/16 \times fr_0$ to fr_0 , in which fr_0 is the frame rate of the original video sequence as well as according to $D+1=5$ resolutions.

[0100] The scrambling of the video sequence is performed for each GOF in the following manner:

- In the temporal subband $t-L_4$ the wavelet coefficients of spatial subbands $s-HH_2$ and $s-HH_3$ resulting from the spatial decomposition into wavelets of frame $LLLL_0$ are extracted and replaced by random or calculated values.

- In the temporal subband $t-H_3$ the wavelet coefficients of spatial subband $s-HH_3$ resulting from the spatial decomposition into wavelets of frame LLH_0 are extracted and replaced by random or calculated values.

- In the temporal subband $t-H_1$ the wavelet coefficients of spatial subband $s-HH_3$ resulting from the spatial decomposition into wavelets of frame LH_0 are extracted and replaced by random or calculated values.

- In the temporal subband $t-H_0$ the wavelet coefficients of spatial subbands $s-HH_3$ resulting from the spatial decomposition into wavelets of frame H_0 are extracted and replaced by random or calculated values.

[0101] The video sequence decoded from the modified main stream is thus totally scrambled unless it is decoded at frame rates equal to $1/16xfr_0$ and $1/8xfr_0$ (decoding solely from respectively temporal subbands $s-LL_0$, $s-LH_1$, $s-HL_1$ and $s-HH_1$ that were not modified). Thus, modifications are made in all the temporal bands but not at all the resolutions. The fact of leaving non-modified resolutions makes it possible to reconstitute the video stream from the scrambled stream, but at a quality that is distinctly less than that of the original video stream.

[0102] The stream scrambled in this manner is transmitted to client 8 upon his request and the descrambling is then performed, e.g., in five stages corresponding to different levels of quality obtained after each descrambling stage. In this manner a descrambling of one or more layers of scalability is carried out and the quality of the film viewed is controlled by the server as a function of the rights of the user and of the quality required by him.

[0103] This descrambling is advantageously expressed by a progressive attenuation of the degradation in time until the reconstitution of the original content with high visual quality. For example, the first descrambling stage consists of restoring the wavelet coefficients of spatial subband $s\text{-HH}_2$ for temporal subband $t\text{-L}_4$. The scrambling of the video sequence decoded for a maximum resolution R and at a maximum frame rate of fr_0 is then less extended spatially and more concentrated around the spatial discontinuities of each frame. The video sequence decoded for the frame rates $1/16 \times fr_0$ or $1/8 \times fr_0$ and for resolutions $r=R/16, R/8, R/4$ is, on the other hand, not scrambled at all. The visual quality of the partially descrambled film is minimal and the film can not be used visually at its full resolution and at its original frame rate. This stage serves to identify the server of complementary information during the establishing of the connection.

[0104] The second descrambling stage consists in restoring the wavelet coefficients of spatial subband $s\text{-HH}_3$ for temporal subband $t\text{-L}_4$. The scrambling of the video sequence decoded for a resolution R and at a maximum frame rate of fr_0 is then less extended spatially and more concentrated around the spatial discontinuities of each frame. Furthermore, the video sequence is now scrambled only for a duration equivalent to one half of a GOF (8 frames of 16). The video sequence decoded for frame rates $1/16 \times fr_0$ or $1/8 \times fr_0$ and for resolutions $r=R/16, R/8, R/4$ is not scrambled. This stage serves to enable user 8 to perceive the video film partially in order to decide whether he wants to obtain the rights to see the film. After the confirmation of the client stages three, four or five are carried out as a function of the payment made. The third descrambling stage consists in restoring the wavelet coefficients of spatial subband $s\text{-HH}_3$ for temporal subband $t\text{-H}_3$. The video sequence is now scrambled only for a duration equivalent to one quarter of a GOF (4 frames out of 16). The video sequence decoded for the frame rates

$1/16xfr_0$, $1/8xfr_0$, $1/4xfr_0$ and for resolutions $r=R/16$, $R/8$, $R/4$ is not scrambled. The film can be viewed visually but it has a low quality.

[0105] The fourth descrambling stage consists in restoring the wavelet coefficients of spatial subband $s\text{-}HH_3$ for temporal subband $t\text{-}H_2$. The video sequence is now scrambled only for a duration equivalent to one eighth of a GOF (2 frames out of 16). The video sequence decoded for frame rates $1/16xfr_0$, $1/8xfr_0$, $1/4xfr_0$, $1/2xfr_0$ and for resolutions $r=R/16$, $R/8$, $R/4$ is not scrambled. The visual quality of the restored film is average. For these 3rd and 4th descrambling stages the video film remains full resolution and remains partially scrambled for the original frame rate but it is possible to extract video streams from these stages with resolutions and frame rates lower than those of the original film. This yields the possibility of supplying versions of the same video stream with a lesser resolution, therefore a lower price and with better control of the access.

[0106] The fifth descrambling stage consists in restoring the wavelet coefficients of spatial subband $s\text{-HH}_3$ for temporal subband $t\text{-H}_1$. The video sequence is now totally descrambled, whatever the frame rate of decoding and the resolution. The reconstituted video stream is strictly identical to the original video stream.